

Direct Variational Perspective Shape from Shading with Cartesian Depth Parametrisation

Yong Chul Ju, Daniel Maurer, Michael Breuß and Andrés Bruhn

Abstract Most of today’s state-of-the-art methods for perspective shape from shading are modelled in terms of partial differential equations (PDEs) of Hamilton-Jacobi type. To improve the robustness of such methods w.r.t. noise and missing data, first approaches have recently been proposed that seek to embed the underlying PDE into a variational framework with data and smoothness term. So far, however, such methods either make use of a radial depth parametrisation that makes the regularisation hard to interpret from a geometrical viewpoint or they consider indirect smoothness terms that require additional consistency constraints to provide valid solutions. Moreover the minimisation of such frameworks is an intricate task, since the underlying energy is typically non-convex. In this chapter we address all three of the aforementioned issues. First, we propose a novel variational model that operates directly on the Cartesian depth. In contrast to existing variational methods for perspective shape from shading this refers to both the data and the smoothness term. Moreover, we employ a direct second-order regulariser with edge-preservation property. This direct regulariser yields by construction valid solutions without requiring

Yong Chul Ju
Institute for Visualization and Interactive Systems
University of Stuttgart, Germany
e-mail: ju@vis.uni-stuttgart.de

Daniel Maurer
Institute for Visualization and Interactive Systems
University of Stuttgart, Germany
e-mail: maurer@vis.uni-stuttgart.de

Michael Breuß
Applied Mathematics and Computer Vision Group
Brandenburg University of Technology Cottbus-Senftenberg, Germany
e-mail: breuss@b-tu.de

Andrés Bruhn
Institute for Visualization and Interactive Systems
University of Stuttgart, Germany
e-mail: bruhn@vis.uni-stuttgart.de

additional consistency constraints. Finally, we also propose a novel coarse-to-fine minimisation framework based on an alternating explicit scheme. This framework allows us to avoid local minima during the minimisation and thus to improve the accuracy of the reconstruction. Experiments show the good quality of our model as well as the usefulness of the proposed numerical scheme.

1 Introduction

Shape from Shading (SfS) is a classic task in computer vision. Given information on light reflectance and illumination in a photographed scene, the aim of SfS is to compute based on the brightness variation the 3D structure of a depicted object from a single input image. SfS has a wide variety of applications, ranging from large scale problems such as astronomy [41] or terrain reconstruction [7] to small scale tasks such as dentistry [2] or endoscopy [34, 51, 52].

Classical Methods. First approaches to SfS go back to 1951 and 1966, respectively, when Van Diggelen [19] and Rindfleisch [41] used SfS techniques to reconstruct the surface of the moon. Later on in the 1970’s, Horn [24] was the first one to tackle the SfS problem by solving a partial differential equation (PDE) approach. In 1981, he and Ikeuchi were also the first ones to model the SfS problem using a variational framework [28]. The most prominent classical variational approach is given by the work of Horn and Brooks [26]. Assuming a simple *orthographic* projection model, a light source at *infinity* as well as a *Lambertian* reflectance model, they proposed to compute the normals of the unknown surface as minimiser of an energy functional.

Those first approaches, however, had several drawbacks. The model assumptions were *very simple* and mainly suitable in the context of astronomical applications. In fact, the use of an orthographic projection model with a light source located at infinity requires the distances between camera, light source and illuminated object to be huge. Also the *depth was not estimated directly* such that the SfS process required a postprocessing step that performed a numerical integration of the estimated surface normals. Thereby, *inconsistent gradient fields* turned out to be a problem, so that extensions of the original model were required that tried to enforce this consistency during or after the estimation [22, 27]. Finally, in case of variational methods, the smoothness term was restricted to a quadratic regulariser [26, 28]. While such standard smoothness terms simplify the minimisation of the underlying energy, they do not allow to preserve discontinuities in the depth and thus lead to *oversmoothed solutions* [35]. For a detailed review of most of the classical methods the reader is referred to [20, 25, 27, 54].

Perspective Shape from Shading. At the end of the 1990’s research mainly focused on novel concepts for formulating orthographic SfS such as viscosity solutions [43] and level set formulations [30]. However, for most applications results were not satisfactory [54]. In the early 2000’s, the situation changed completely. Inspired by the work of Okatani and Deguchi [34], independently, Prados and his co-workers

[38, 39] as well as two other research groups [16, 46] proposed to consider a *perspective* camera model. Evidently, such a model is particularly appropriate for tasks that require the object to be relatively close to the camera such as e.g. in medical endoscopy. In such cases the perspective effects dominate and an orthographic projection model would cause significant systematic errors as shown in [47]. Secondly, Prados and colleagues proposed to *shift the light source location from infinity to the camera centre* which can be seen as a good approximation of a camera with photoflash. This made shape from shading attractive for a variety of photo-based applications. Finally, also a physically motivated *light attenuation term* was introduced that models a quadratic fall-off due to the inverse square law. As discussed in [9], the use of this term largely resolved the convex-concave ambiguity that was inherent to the classical orthographic model although some ambiguities are still present. Even the generalisation of such approaches to advanced reflectance models such as the Oren-Nayar [36] or the Phong reflectance model [37] have been recently investigated [4, 48].

However, this evolution of SfS models was accompanied by a different way of formulating the SfS problem. Instead of using variational methods, the perspective SfS problem was formulated in terms of hyperbolic PDEs [39]. Although such PDE formulations allow for an efficient computation of the solution using fast marching schemes [45], they suffer from two inherent drawbacks: (i) On the one hand, they are prone to *noise* and *missing data*, since they do not rely on any form of regularisation or filling-in. This can be particularly problematic in the context of real-world images. (ii) On the other hand, it is *difficult to extend* the underlying model of such PDE-based schemes by additional constraints such as smoothness terms, multiple views, or additional light sources. While there have been recently some PDE-based approaches to photometric stereo [32], one has to take care of ensuring the uniqueness of the solution if the input data from multiple images is not consistent, cf. the discussion in [33].

Variational Perspective Shape from Shading. Given the flexibility and robustness of variational methods, it is not surprising that recently researchers tried to close the evolutionary loop by integrating the perspective SfS model into a suitable variational framework. So far, however, there are only a few works in the literature that deal with this recent idea. On the one hand, there is the work of Ju *et al.* [29] that embeds the PDE of Prados *et al.* [39] as data term into a variational model and complements it with a discontinuity-preserving second order smoothness term. However, since the approach penalises deviations from the PDE directly and uses a parametrisation in terms of the radial depth, deviations in both the data and the smoothness term are difficult to interpret geometrically or photometrically. On the other hand, there is the approach of Abdelrahim *et al.* [3] that formulates the data term in terms of brightness differences and makes use of a Cartesian depth parametrisation. While the corresponding energy functional is thus more meaningful from a geometric and photometric viewpoint, it defines smoothness based on surface normals and thus needs an additional integrability constraint. Moreover, the corresponding smoothness term is restricted to a simple homogeneous regulariser that does not allow to preserve object edges during the reconstruction. Finally, there are

the works of Zhang *et al.* [55] and Wu *et al.* [52] that also make use of a Cartesian depth parametrisation but rely on an indirect estimation using auxiliary variables. While the approach of Zhang *et al.* [55] resolves the resulting consistency problem by considering an integrability constraint, the method of Wu *et al.* [52] repeatedly integrates the surface normals during computation to ensure valid solutions. Moreover, both approaches use derivations for their surface normals that are based on the orthographic projection model of Horn and Brooks [27]. Unfortunately, the resulting models are thus only valid in case of weak perspective distortions.

A final issue that is common to all four of the aforementioned works is the difficulty of minimising the underlying energy. Since this energy is non-convex, two of the methods rely on initialisations provided by closely related PDE-based SfS approaches [3, 55]. This, however, contradicts the idea of introducing robustness into the estimation – in particular in the presence of noise or missing data. In contrast, the other two methods estimate the solution from scratch [29, 52]. However, those methods do not provide any quantitative assessment of the reconstruction quality.

Let us summarise: While from a modelling viewpoint, it would be desirable to design a variational model that directly solves for the Cartesian depth without the need of integrability constraints or repeated integrations steps, it would be helpful from an optimisation viewpoint to develop a minimisation scheme that neither depends on the solution of other SfS techniques as in [3, 55] nor requires an accurate initialisation to produce meaningful results.

Our Contributions. In this book chapter we contribute to the field of variational SfS in three ways: (i) First, we consider a variational model for perspective SfS that makes use of a Cartesian depth parametrisation and an edge-preserving Cartesian depth regularisation. By penalising deviations from the image brightness in the data term and regularising the Cartesian depth in the smoothness term directly, we obtain an approach that is geometrically and photometrically meaningful. In this context, we also point out a popular mistake in the derivation of the surface normal and show two different ways to derive the normal correctly. (ii) Our method is a direct approach to depth computation, i.e. it does not yield gradient fields that need to be integrated in a subsequent step, nor do we employ integrability constraints. (iii) Apart from the novel model, we also propose a novel minimisation strategy. By embedding an alternating explicit scheme into a coarse-to-fine scheme, we obtain an optimisation framework that allows to obtain significantly better results than a traditional explicit scheme. Experiments with synthetic and real-world images show the good quality of our reconstructions and the advantages of our numerical scheme.

Organisation of the Chapter. In Section 2 we propose a novel PDE-based model for perspective SfS that is based on a Cartesian parametrisation of the depth. In Section 3 we then embed this PDE into a variational framework with appropriate second order smoothness term. Details on the minimisation and the discretisation are provided in Section 4, while Section 5 comments on the integration of intrinsic camera parameters. Finally, a detailed evaluation of our approach is presented in Section 6. The paper concludes with a summary in Section 7.

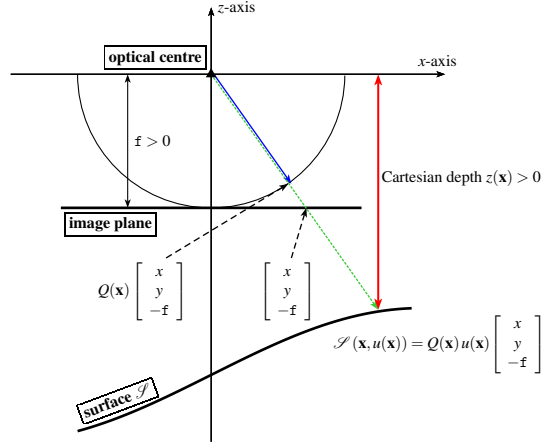


Fig. 1 Relation between the radial depth factor $u(\mathbf{x})$ (quotient between green and blue distance) that denotes the depth in multiples of the focal length f and the Cartesian depth $z(\mathbf{x})$ (red distance).

2 Perspective SfS with Cartesian Depth Parametrisation

In this section, we introduce a novel PDE-based SfS model that is parametrised in terms of the *Cartesian* depth. In contrast to most existing SfS models that estimate the radial depth or multiples thereof, such a Cartesian parametrisation expresses the unknown surface directly in terms of the Euclidean distance along the z -axis, which is the axis orthogonal to the image plane.

Parametrisation of the Surface. The starting point for our new model is formed by the classical PDE-approach of Prados *et al.* [39] which is originally parametrised in terms of the *radial* depth. Key assumptions of this SfS model are that a point light source is located at the *optical centre* of a perspective camera and that the surface reflectance is *Lambertian* with uniform albedo that is fixed to one. The unknown surface $\mathcal{S} : \Omega_{\mathbf{x}} \rightarrow \mathbb{R}^3$ can then be described as

$$\mathcal{S}(\mathbf{x}, u(\mathbf{x})) = \left\{ \frac{f}{\sqrt{|\mathbf{x}|^2 + f^2}} u(\mathbf{x}) \begin{bmatrix} x \\ y \\ -f \end{bmatrix} \mid \mathbf{x} := (x, y)^\top \in \overline{\Omega_{\mathbf{x}}} \right\}, \quad (1)$$

where $\mathbf{x} = (x, y)^\top \in \overline{\Omega_{\mathbf{x}}}$ is the position in the closure $\overline{\Omega_{\mathbf{x}}}$ of the rectangular image domain $\Omega_{\mathbf{x}} \subset \mathbb{R}^2$, f denotes the focal length of the camera and $u(\mathbf{x})$ is a multiple of f that describes the radial distance (depth) of the surface from the camera centre.

Since the third component in Eq. (1) corresponds to the negative Cartesian depth z , we can derive the following relationship to the radial depth $u f$

$$z(\mathbf{x}) = \frac{f}{\sqrt{|\mathbf{x}|^2 + f^2}} u(\mathbf{x}) f \stackrel{(1)}{=} Q(\mathbf{x}) u(\mathbf{x}) f, \quad (2)$$

where $Q(\mathbf{x})$ denotes a spatially variant conversion factor given by

$$Q(\mathbf{x}) = \frac{f}{\sqrt{|\mathbf{x}|^2 + f^2}}. \quad (3)$$

This relation is illustrated in Figure 1.

Plugging Eq. (3) into Eq. (1), we then obtain the parametrisation of the original surface \mathcal{S} with respect to the Cartesian depth z

$$\mathcal{S}(\mathbf{x}, z(\mathbf{x})) \stackrel{(1)}{=} Q(\mathbf{x}) u(\mathbf{x}) \begin{bmatrix} x \\ y \\ -f \end{bmatrix} \stackrel{(2)}{=} Q(\mathbf{x}) \left(\frac{z(\mathbf{x})}{f Q(\mathbf{x})} \right) \begin{bmatrix} x \\ y \\ -f \end{bmatrix} = \begin{bmatrix} \frac{z(\mathbf{x})x}{f} \\ \frac{z(\mathbf{x})y}{f} \\ -z(\mathbf{x}) \end{bmatrix}. \quad (4)$$

Brightness Equation. After we have parametrised the original surface in terms of the Cartesian depth, let us now derive the resulting brightness equation that relates the local orientation of the surface to the image brightness. Assuming a Lambertian reflectance model and a quadratic light attenuation term that follows the inverse square law, we obtain the following general brightness equation [39]:

$$I(\mathbf{x}) = \frac{1}{r(\mathbf{x})^2} \left(\frac{\mathbf{n}(\mathbf{x})}{|\mathbf{n}(\mathbf{x})|} \cdot \mathbf{L}(\mathbf{x}) \right), \quad (5)$$

where I is the recorded image, \mathbf{n} is the surface normal vector, \mathbf{L} stands for the normalised light direction vector, and r is the (radial) distance of the light source to the surface. Knowing that $r = fu$ and using Eq. (2) we can express the quadratic light attenuation term using the Cartesian depth z

$$r(\mathbf{x}) = fu(\mathbf{x}) = \frac{z(\mathbf{x})}{Q(\mathbf{x})} \Rightarrow \frac{1}{r(\mathbf{x})^2} = \frac{Q(\mathbf{x})^2}{z(\mathbf{x})^2}. \quad (6)$$

What remains to be computed in terms of the Cartesian depth are the surface normal \mathbf{n} and the light direction vector \mathbf{L} , respectively.

Surface Normal. Let us start by deriving the surface normal. Since the surface normal is the normal vector of the tangent plane, we first have to compute the partial derivatives of the surface in Eq. (4) in x - and y -direction, respectively

$$\mathcal{S}_x(\mathbf{x}, z) = \begin{bmatrix} \frac{z_x x + z}{f} \\ \frac{z_x y}{f} \\ -z_x \end{bmatrix}, \quad \mathcal{S}_y(\mathbf{x}, z) = \begin{bmatrix} \frac{z_y x}{f} \\ \frac{z_y y + z}{f} \\ -z_y \end{bmatrix}. \quad (7)$$

Here and for the whole paper we dropped the spatial dependency of z , z_x and z_y on \mathbf{x} for the sake of clarity. Taking the cross-product then yields the direction of the surface normal

$$\mathbf{n}(\mathbf{x}) = \mathcal{S}_x(\mathbf{x}, z) \times \mathcal{S}_y(\mathbf{x}, z) = \begin{bmatrix} \frac{z_x z}{f} \\ \frac{z_y z}{f} \\ \frac{z [(\nabla z \cdot \mathbf{x}) + z]}{f^2} \end{bmatrix}. \quad (8)$$

Light Direction. Let us now turn towards the computation of the light direction. Since the light source is assumed to be located in the camera centre which coincides with the origin of the coordinate system, the direction of the light rays and the direction of the optical rays coincide (up to sign). Hence, the light direction can just be read off Eq. (1) as

$$\mathbf{L}(\mathbf{x}) = \frac{1}{\sqrt{|\mathbf{x}|^2 + f^2}} \begin{bmatrix} -x \\ -y \\ f \end{bmatrix}. \quad (9)$$

PDE-Based Model. By plugging the surface normal from Eq. (8) and the light direction from Eq. (9) into the brightness equation (5) we finally obtain our perspective SfS model with the new Cartesian depth parametrisation

$$I - \frac{Q^3}{z \sqrt{f^2 |\nabla z|^2 + [(\nabla z \cdot \mathbf{x}) + z]^2}} = 0. \quad (10)$$

Here and for the whole paper we dropped the spatial dependency of I and Q on \mathbf{x} for the sake of clarity.

The main properties of our new model (10) are naturally inherited from the original PDE [39]: (i) Eq. (10) still belongs to the class of Hamilton-Jacobi equations (HJEs) which have been intensively studied in the SfS literature. (ii) Therefore, well-posedness can be achieved in the viscosity sense [9, 14, 39]. (iii) Proper numerical discretisations must be considered when solving the HJE.

Let us note that the framework of viscosity solutions is a natural setting for HJEs such as Eq. (10). The basic idea behind the notion of viscosity solutions is to add a (typically, second order) regularisation term to the PDE and study the solution as this term goes to zero. This proceeding yields desirable stability properties and enables to consider even solutions with non-differentiable features like e.g. kinks. We refer the interested reader to [5, 14] for studying properties of viscosity solutions and to [12] for their use in computer vision.

Furthermore, please note that our model can be seen as a generalisation of the PDE-based approach in [56] that already makes use of the Cartesian depth parametrisation, but does not yet consider the light attenuation term from physics.

3 Variational Model for Perspective SfS with Cartesian Depth Parametrisation

So far we have derived a novel PDE-based model for perspective SfS with Cartesian depth parametrisation. Let us now discuss how this model can be integrated into a variational framework with smoothness term.

Variational Model. To this end, we follow the idea from [29] and use a quadratic error term based on our novel PDE as data term which is complemented with a suitable second order regulariser. More precisely, we propose to compute the Cartesian depth z as minimiser of the following energy functional

$$E(z) = \int_{\Omega_x} c(\mathbf{x}) \underbrace{D(\mathbf{x}, z, \nabla z)}_{\text{Data term}} + \alpha \underbrace{S(\text{Hess}(z))}_{\text{Smoothness term}} d\mathbf{x}, \quad (11)$$

where D is the data term, S is the smoothness term, $c : \mathbf{x} \in \overline{\Omega_x} \subset \mathbb{R}^2 \rightarrow [0, 1]$ is a confidence function and $\alpha \in \mathbb{R}^+$ is a regularisation parameter that steers the degree of smoothness of the solution. As mentioned before our data term is based on a quadratic formulation that penalises deviations from our novel PDE. It is given by

$$D(\mathbf{x}, z, \nabla z) = \left(I(\mathbf{x}) - \frac{Q(\mathbf{x})^3}{z W(\mathbf{x}, z, \nabla z)} \right)^2 \quad (12)$$

with

$$W(\mathbf{x}, z, \nabla z) = \sqrt{f^2 |\nabla z|^2 + [(\nabla z \cdot \mathbf{x}) + z]^2}. \quad (13)$$

As smoothness term, we propose to use the following subquadratic and thus edge-preserving second-order regulariser based on the Frobenius norm of the Hessian

$$S(\text{Hess}(z)) = \Psi \left(\|\text{Hess}(z)\|_F^2 \right) = \Psi \left(z_{xx}^2 + 2z_{xy}^2 + z_{yy}^2 \right) \quad (14)$$

where Ψ is the Charbonnier function [13]

$$\Psi(s^2) = 2\lambda^2 \sqrt{1 + \frac{s^2}{\lambda^2}} \quad (15)$$

with contrast parameter λ . Such higher-order smoothness terms have already been successfully applied in the context of perspective SfS parametrised in terms of the radial depth [29], orthographic SfS [49], image denoising [31], optical lithography [21] and motion estimation [18]. Finally, the use of the confidence function c in the data term allows to exclude unreliable image regions which have been identified a priori, e.g. by a texture detector or by a background segmentation algorithm. Such functions are particularly useful in the context of real-world images that contain texture, noise, or missing data [17, 29].

Table 1 Comparison of the literature on variational models for perspective shape from shading.

	Zhang <i>et al.</i> [55]	Wu <i>et al.</i> [52]	Abdelrahim <i>et al.</i> [3]	Ju <i>et al.</i> [29]	Our Work
Parametrisation	Cartesian depth	Cartesian depth	Cartesian depth	radial depth	Cartesian depth
Reprojection Error as Data Term	✓	✓	✓	—	✓
Light Attenuation Factor	—	✓	✓ ¹	✓	✓
Correct Surface Normal	— ²	— ²	✓ ³	✓	✓
Regularisation	Cartesian depth	Cartesian depth	Cartesian surface normal	radial depth	Cartesian depth
Edge Preservation	—	—	—	✓	✓
No Integrability Term	—	— ⁴	—	✓	✓
Direct Estimation ⁵	—	—	✓	✓	✓

¹ factor not expressed in terms of the Cartesian depth² see explanation in appendix³ no details given in the paper but derivations shown in [1]⁴ integrability constraint realised via repeated integration of surface normals⁵ depth is computed without extra variables for surface normals

Properties. Our variational model from Eq. (11) has the following distinct features:

- (i) Since the data term in Eq. (12) is inherited from Eq. (10), the perspective camera projection is already taken into account. Moreover, since the reprojection error is penalised in the data term, deviations have a *photometric* interpretation.
- (ii) Since the regulariser is applied directly to the Cartesian depth, also deviations from smoothness become now more meaningful than in the case of a radial depth parametrisation. In particular, they can be interpreted *geometrically*.
- (iii) Moreover, in contrast to most existing approaches, the regulariser is able to *preserve edges* in the reconstruction despite of the regularisation effect.
- (iv) Unreliable regions can be excluded from the data term via a *confidence function* such that the smoothness term takes over and fills in information from the neighbourhood. This can be advantageous in the context of texture, noise, or missing data. Please note that in contrast to [29], we always guarantee a fixed amount of regularisation by not restricting the smoothness term to unreliable locations.
- (v) The depth of the surface is *directly* computed since we minimise for the unknown depth z in Eq. (11). This is in contrast to most variational methods that estimate the depth in two steps, see e.g. [10, 22, 28] where first the surface normals are computed by a variational model and then the depth is determined by integration.

- (vi) The solution given by the model fulfils the *integrability constraint* per construction since we solve for z and use $z_{xy} = z_{yx}$ in the smoothness term. Otherwise, such as in [3], an additional integrability term would be needed to encourage valid solutions.
- (vii) Another advantage of the new parametrisation is that it allows a straightforward *combination* with other reconstruction methods such as stereo [42] or scene flow estimation [6], since such approaches typically make use of the same Cartesian parametrisation and thus could be easily integrated into a joint framework.
- (viii) A final advantage is the fact that the approach could easily be extended to *multiple views*, since transformations between the views are simpler if the approach is parametrised in terms of the Cartesian depth instead of the radial depth.

To make the difference of our model to other variational approaches from the literature explicit, the features of the different methods are compared in Table 1.

4 Minimisation

Let us now discuss the minimisation of the proposed energy. To this end, we will first derive the associated Euler-Lagrange equation and then discuss its discretisation. Finally, we will sketch a coarse-to-fine minimisation strategy with an alternating explicit scheme to solve the resulting nonlinear equations.

Euler-Lagrange Equation. The calculus of variations [15] tells us that the minimiser z of our energy in Eq. (11) has to fulfil the corresponding Euler-Lagrange equation. Omitting the dependencies on all variables in order to ease the readability, this equation is given by

$$\begin{aligned}
0 &= [cD + \alpha S]_z - \frac{\partial}{\partial x} [cD + \alpha S]_{z_x} - \frac{\partial}{\partial y} [cD + \alpha S]_{z_y} \\
&\quad + \frac{\partial^2}{\partial x^2} [cD + \alpha S]_{z_{xx}} + 2 \frac{\partial^2}{\partial x \partial y} [cD + \alpha S]_{z_{xy}} + \frac{\partial^2}{\partial y^2} [cD + \alpha S]_{z_{yy}} \\
&= [cD]_z - \frac{\partial}{\partial x} [cD]_{z_x} - \frac{\partial}{\partial y} [cD]_{z_y} + \frac{\partial^2}{\partial x^2} \overbrace{[cD]_{z_{xx}}}^{\equiv 0} + 2 \frac{\partial^2}{\partial x \partial y} \overbrace{[cD]_{z_{xy}}}^{\equiv 0} + \frac{\partial^2}{\partial y^2} \overbrace{[cD]_{z_{yy}}}^{\equiv 0} \\
&\quad + \underbrace{[\alpha S]_z}_{\equiv 0} - \frac{\partial}{\partial x} \underbrace{[\alpha S]_{z_x}}_{\equiv 0} - \frac{\partial}{\partial y} \underbrace{[\alpha S]_{z_y}}_{\equiv 0} + \frac{\partial^2}{\partial x^2} [\alpha S]_{z_{xx}} + 2 \frac{\partial^2}{\partial x \partial y} [\alpha S]_{z_{xy}} + \frac{\partial^2}{\partial y^2} [\alpha S]_{z_{yy}} \\
&= c \left([D]_z - \frac{\partial}{\partial x} [D]_{z_x} - \frac{\partial}{\partial y} [D]_{z_y} \right) + \frac{\partial^2}{\partial x^2} [\alpha S]_{z_{xx}} + 2 \frac{\partial^2}{\partial x \partial y} [\alpha S]_{z_{xy}} + \frac{\partial^2}{\partial y^2} [\alpha S]_{z_{yy}},
\end{aligned} \tag{16}$$

where we exploited the fact that

$$\frac{\partial^2}{\partial x \partial y} [cD + \alpha S]_{z_{xy}} = \frac{\partial^2}{\partial y \partial x} [cD + \alpha S]_{z_{xy}}. \tag{17}$$

On a structural level, this Euler-Lagrange equation is somewhat more complicated than its counterparts for indirect methods in [52, 55]. Such indirect methods model the surface normal using auxiliary variables $p = z_x$ and $q = z_y$ and thus do not have the additional data term contributions $\frac{\partial}{\partial x} [D]_{z_x}$ and $\frac{\partial}{\partial y} [D]_{z_y}$.

Let us now take a closer look at all the individual terms that occur in Eq. (17). After some computations we obtain

$$\begin{aligned} [D]_z &= 2 \left(I - \frac{Q^3}{zW} \right) \left(\frac{Q^3}{z^2 W} + \frac{Q^3}{zW^2} [W]_z \right) \\ &= 2 \left(I - \frac{Q^3}{zW} \right) \frac{Q^3}{zW} \left(\frac{1}{z} + \frac{\nabla z \cdot \mathbf{x} + z}{W^2} \right) \end{aligned} \quad (18)$$

$$\begin{aligned} \frac{\partial}{\partial x} [D]_{z_x} &= \left[2 \left(I - \frac{Q^3}{zW} \right) \frac{Q^3}{zW^3} [W]_x \right]_x \\ &= \left[2 \left(I - \frac{Q^3}{zW} \right) \frac{Q^3}{zW^3} (\mathbf{f}^2 z_x + [\nabla z \cdot \mathbf{x} + z] x) \right]_x \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{\partial}{\partial y} [D]_{z_y} &= \left[2 \left(I - \frac{Q^3}{zW} \right) \frac{Q^3}{zW^3} [W]_y \right]_y \\ &= \left[2 \left(I - \frac{Q^3}{zW} \right) \frac{Q^3}{zW^3} (\mathbf{f}^2 z_y + [\nabla z \cdot \mathbf{x} + z] y) \right]_y \end{aligned} \quad (20)$$

as well as

$$\frac{\partial^2}{\partial x^2} [S]_{z_{xx}} = 2 \frac{\partial^2}{\partial x^2} \left[\Psi'(\|\text{Hess}(z)\|_F^2) z_{xx} \right], \quad (21)$$

$$2 \frac{\partial^2}{\partial xy} [S]_{z_{xy}} = 4 \frac{\partial^2}{\partial xy} \left[\Psi'(\|\text{Hess}(z)\|_F^2) z_{xy} \right], \quad (22)$$

$$\frac{\partial^2}{\partial y^2} [S]_{z_{yy}} = 2 \frac{\partial^2}{\partial y^2} \left[\Psi'(\|\text{Hess}(z)\|_F^2) z_{yy} \right], \quad (23)$$

where the derivative of the penaliser function $\Psi(s^2)$ reads

$$\Psi'(s^2) = \frac{\partial}{\partial(s^2)} \Psi(s^2) = \frac{1}{\sqrt{1 + \frac{s^2}{\lambda^2}}}. \quad (24)$$

While the contributions of the data term are related to the influence of z and ∇z on the brightness equation, the contributions of the smoothness term define an edge-preserving fourth-order diffusion process. This becomes explicit as follows: Since $\Psi'(s^2)$ becomes small for large values of s^2 , this reduces the effect of the smoothing at locations with high curvature, i.e. where $\|\text{Hess}(z)\|_F^2$ is large. After we have

derived the resulting Euler-Lagrange equation, let us now discuss how this equation can be discretised appropriately.

Discretisation. In order to discretise the contributions of the data term given by Eqs. (18) – (20), we employ the upwind scheme from [43] in view of the hyperbolic nature of the underlying PDE. In 1D, the corresponding upwind discretisation reads

$$\tilde{z}_x \approx \max(D^-z, -D^+z, 0), \quad (25)$$

with

$$D^-z = \frac{z_i - z_{i-1}}{h_x} \quad \text{and} \quad D^+z = \frac{z_{i+1} - z_i}{h_x}, \quad (26)$$

where h_x denotes the grid size. Please note that in contrast to upwind schemes for eikonal equations [45] that typically approximate only the magnitude of the gradient, the sign matters in our case, such that we have to choose

$$z_x = \begin{cases} D^+z & \text{if } \tilde{z}_x = -D^+z, \\ \tilde{z}_x & \text{otherwise.} \end{cases} \quad (27)$$

This selects the actual forward difference, if the second argument in (25) is the maximum [8, 9]. This scheme can be extended in a straightforward way to 2D. For discretising the contributions of the smoothness term, a standard central difference scheme is used.

Since it is difficult to discretise the Euler-Lagrange equation directly, we followed a *first discretise then optimise* scheme. To this end, we used the aforementioned finite difference approximations to discretise the energy in (11) applying the upwind scheme for the data term and a central difference approximation for the smoothness term. Then, by computing the derivatives of the discrete energy we obtain a proper discretisation for the Euler-Lagrange equation.

Finally, by using the Euler forward time discretisation method

$$z_t \approx \frac{z^{n+1} - z^n}{\tau}, \quad (28)$$

with τ being a time step size, we can reformulate the solution of Eq. (17) as the steady state of the corresponding evolution equation in artificial time. Thus we obtain the following explicit scheme

$$\frac{z^{n+1} - z^n}{\tau} + EL^n = 0 \quad \Leftrightarrow \quad z^{n+1} = z^n - \tau EL^n, \quad (29)$$

where EL^n is the discretisation of the Euler-Lagrange equation evaluated at time n . Please note that this discretisation may change over time, since we re-discretised the energy in each iteration by adapting the direction of the discretisation of the upwind scheme (forward, backward, no contribution) based on evaluating Eqs. (25) – (27) for the result of the previous time step. In that sense we use a *lagged* discretisation approach, where the discretisation is updated in each iteration.

Coarse-to-Fine Approach. Since the underlying energy functional is highly non-convex, the proposed explicit scheme may get trapped in local minima. To tackle this problem, we propose to embed the estimation into a coarse-to-fine framework. Starting from a very coarse resolution, we successively refine the input image while repeatedly reconstructing the surface. Thereby, solutions from coarser levels serve as initialisation for the finer scales. Similar hierarchical schemes have already been successfully applied to many other problems in computer vision; see e.g. [8, 11].

Apart from improving the quality of the results by avoiding local minima, coarse-to-fine schemes also render the estimation more robust w.r.t. the choice of the initialisation. In fact, if sufficiently many resolution levels were used, we could hardly observe any impact of the initialisation on the quality of the final results. Since a good initial guess can still be useful to speed up the computation, we propose to initialise the depth by pointwise solving the data term in Eq. (12) for $\nabla z = 0$

$$D(\mathbf{x}, z, 0) = 0 \quad \Rightarrow \quad z = \sqrt{\frac{Q(\mathbf{x})^3}{I(\mathbf{x})}}. \quad (30)$$

This can be seen as an efficient compromise between using the full model which is evidently not feasible and only considering the inverse square law, i.e. $z = 1/\sqrt{I(\mathbf{x})}$, which completely neglects the effect of the surface orientation and thus actually provides a local upper bound for the correct depth. In any case, in contrast to other variational SfS methods from the literature, our technique does not have to rely on initialisations from non-variational SfS approaches [3, 55] or surface integration methods [52] to provide meaningful results.

Let us now discuss the details of our coarse-to-fine approach. To this end, we introduce the parameter η that specifies the downsampling factor between two consecutive resolution levels and that is typically chosen in the interval $(0.5, 1)$. Then the grid size at level k of our coarse-to-fine approach can be computed as

$$h_x^k = h_x \cdot \eta^{-k}, \quad h_y^k = h_y \cdot \eta^{-k}. \quad (31)$$

where $k = 0$ is the original resolution and $k = k_{\max}$ is the coarsest level. This tells us that the grid size becomes larger at coarser scales which intuitively makes sense, since the size of the image plane remains constant while the number of pixels decreases. At the same time, however, this increase of the grid size leads to a major problem: Since the contributions of the smoothness term given by Eqs. (21) – (23) involve fourth-order derivatives that scale proportionally to $1/h^4$, the strength of the regularisation actually decreases with η^{4k} on coarser scales. In order to compensate for this effect, we thus propose to scale the smoothness weight α according to

$$\alpha^k = \eta^{-4k} \cdot \alpha. \quad (32)$$

This guarantees a similar amount of regularisation for all resolution levels.

Alternating Explicit Scheme. Finally, we observed in our experiments that the terms in Eqs. (19) and (20) that refer to the influence of the depth gradient ∇z on

the brightness equation require to select the time step size τ rather small. In particular, these terms do not have a weighting parameter such as the smoothness term that can be adjusted appropriately. As a consequence, the minimisation typically needs several thousands or even millions of iterations. To counter this problem, we propose the following alternating estimation strategy at each resolution level: For a fixed number of iterations n , instead of performing n iterations using the original explicit scheme, we propose an alternating iterative scheme that first does $n/2$ iterations with a simplified explicit scheme neglecting the two terms in Eqs. (19) and (20), followed by $n/2$ iterations with the entire explicit scheme. Since the neglected terms are based on second-order derivatives and the remaining terms did not strongly affect the convergence, we empirically found out that we can choose the time step size approximately $\min(h_x^{-2}, h_y^{-2})$ times larger for the first $n/2$ iterations (given that $h_x, h_y \ll 1$). In our experiments this leads to speed-ups of about one to four orders of magnitude. Moreover, in most cases, even the simplified scheme was sufficient to achieve excellent results. Thereby one should note that, from a numerical viewpoint, the simplified scheme can be understood as an optimisation method for a series of energy functionals of type of Eq. (11), where the gradient ∇z is lagging and thus has no direct influence on the optimisation.

5 Intrinsic Parameters

So far we have derived a variational model for perspective SfS with Cartesian depth parametrisation that is given in terms of *image coordinates*. Let us now discuss how the model and the minimisation has to be adapted if we additionally consider the intrinsic camera parameters, i.e. if we express the model in terms of *pixel coordinates*.

Coordinate Transformation. Let the corresponding calibration matrix be given by

$$K = \begin{bmatrix} \mathbf{f}/h_x & 0 & c_1 \\ 0 & \mathbf{f}/h_y & c_2 \\ 0 & 0 & 1 \end{bmatrix}. \quad (33)$$

where $(c_1, c_2)^\top$ denotes the location of the focal point, and h_x and h_y is the grid size in x - and y -direction, respectively [23]. Knowing this matrix allows us to reformulate the image coordinates $\mathbf{x} = (x, y)^\top$ of our original model in terms of pixel coordinates $\mathbf{a} = (a, b)^\top$. The corresponding transformation is given by

$$\begin{bmatrix} a \\ b \\ -1 \end{bmatrix} = K \frac{1}{\mathbf{f}} \begin{bmatrix} x \\ y \\ -\mathbf{f} \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \\ -\mathbf{f} \end{bmatrix} = \mathbf{f} K^{-1} \begin{bmatrix} a \\ b \\ -1 \end{bmatrix}, \quad (34)$$

where one has to take care that the image plane is at distance \mathbf{f} of the camera centre. Plugging Eq. (33) into Eq. (34) then yields

$$\mathbf{x}(\mathbf{a}) = \begin{bmatrix} x(a) \\ y(b) \end{bmatrix} = \begin{bmatrix} h_x & 0 \\ 0 & h_y \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} h_x c_1 \\ h_y c_2 \end{bmatrix}. \quad (35)$$

Variational Model. Now we are in the position to reformulate our entire model in terms of pixel coordinates. Substituting Eq. (35) into our original energy and transforming the integration domain $\bar{\Omega}_{\mathbf{a}} = \mathbf{x}^{-1}(\bar{\Omega}_{\mathbf{x}})$ accordingly, we obtain the following variational model expressed in terms of pixel coordinates

$$E(z(\mathbf{x}(\mathbf{a}))) = \int_{\bar{\Omega}_{\mathbf{a}}} c(\mathbf{x}(\mathbf{a})) \underbrace{D(\mathbf{x}(\mathbf{a}), z(\mathbf{x}(\mathbf{a})), \nabla z(\mathbf{x}(\mathbf{a})))}_{\text{Data term}} + \alpha \underbrace{S(\text{Hess}(z)(\mathbf{x}(\mathbf{a})))}_{\text{Smoothness term}} d\mathbf{a}. \quad (36)$$

Please note that we omitted the substitution factor given by $|\det(J(\mathbf{x}(\mathbf{a})))|$, where J is the Jacobian, since this factor is constant and thus does not change the minimiser of our energy. Let us now derive the corresponding Euler-Lagrange equation for our novel model expressed in terms of pixel coordinates.

Euler-Lagrange Equation. Analogously to Eq. (17) we drop the dependencies on all variables and obtain the following Euler-Lagrange equation

$$\begin{aligned} 0 &= c \left([D]_z - \frac{\partial}{\partial a} [D]_{z_a} - \frac{\partial}{\partial b} [D]_{z_b} \right) \\ &\quad + \alpha \left(\frac{\partial^2}{\partial a^2} [S]_{z_{aa}} + 2 \frac{\partial^2}{\partial a \partial b} [S]_{z_{ab}} + \frac{\partial^2}{\partial b^2} [S]_{z_{bb}} \right) \end{aligned} \quad (37)$$

$$\begin{aligned} &= c \left([D]_z - \frac{\partial}{\partial x} [D]_{z_x} - \frac{\partial}{\partial y} [D]_{z_y} \right) \\ &\quad + \alpha \left(\frac{\partial^2}{\partial x^2} [S]_{z_{xx}} + 2 \frac{\partial^2}{\partial x \partial y} [S]_{z_{xy}} + \frac{\partial^2}{\partial y^2} [S]_{z_{yy}} \right), \end{aligned} \quad (38)$$

where we exploited the following relation between derivatives in pixel and image coordinates due to Eq. (35)

$$\frac{\partial}{\partial a} = h_x \frac{\partial}{\partial x}, \quad \frac{\partial}{\partial b} = h_y \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial z_a} = \frac{1}{h_x} \frac{\partial}{\partial z_x}, \quad \frac{\partial}{\partial z_b} = \frac{1}{h_y} \frac{\partial}{\partial z_y}, \quad (39)$$

$$\frac{\partial}{\partial z_{aa}} = \frac{1}{h_x h_x} \frac{\partial}{\partial z_{xx}}, \quad \frac{\partial}{\partial z_{ab}} = \frac{1}{h_x h_y} \frac{\partial}{\partial z_{xy}}, \quad \frac{\partial}{\partial z_{bb}} = \frac{1}{h_y h_y} \frac{\partial}{\partial z_{yy}}. \quad (40)$$

The equality between Eq. (37) and Eq. (38) shows that the Euler-Lagrange equations of our models in pixel and image coordinates are basically identical. One only has to parametrise the terms (18)–(23) that have been originally derived in image coordinates using the coordinate transform in Eq. (35). Apart from that, the discretisation can be performed in accordance with our explanations from the previous section. In this context, the grid size is given by the intrinsic parameters h_x and h_y . Moreover, one has to adapt the camera matrix K at each level of the coarse-to-fine scheme. This requires to scale both the grid size and the principal point $(c_1, c_2)^\top$.

6 Evaluation

Test Images and Error Measures. In order to evaluate our novel approach we make use of four synthetic images with ground truth that fulfil the underlying assumptions regarding reflectance and illumination. This allows us to compute two error measures: one with respect to the reconstructed surface and the other one with respect to reprojected image. The first error measure is the *relative surface error* (RSE) of a point wise computed Euclidean distance between the computed surface \mathcal{S} and the ground truth surface \mathcal{S}^{gt} . It is given by

$$\text{RSE} = \frac{\sum_{\bar{\Omega}_a} |\mathcal{S}(\mathbf{x}(\mathbf{a})) - \mathcal{S}^{\text{gt}}(\mathbf{x}(\mathbf{a}))|}{\sum_{\bar{\Omega}_a} |\mathcal{S}^{\text{gt}}(\mathbf{x}(\mathbf{a}))|}, \quad (41)$$

where the normalisation allows to determine the reconstruction error *relative* to the ground truth shape. This in turn makes errors of differently scaled surfaces comparable. The second error measure is the *relative image error* (RIE) between the reprojected image I and the given input image I^{gt} . It is defined as follows

$$\text{RIE} = \frac{\sum_{\bar{\Omega}_a} |I(\mathbf{x}(\mathbf{a})) - I^{\text{gt}}(\mathbf{x}(\mathbf{a}))|}{\sum_{\bar{\Omega}_a} |I^{\text{gt}}(\mathbf{x}(\mathbf{a}))|}. \quad (42)$$

This time, however, the normalisation is performed with respect to the brightness of the input image to make reprojection results for input images with different brightness scale comparable. Summarising: While the first measure reflects how well the reconstruction matches the ground truth surface, the second measure determines how well the reprojection fits the input data.

Let us now discuss the considered test images which are depicted in Fig. 2 in detail. The first synthetic test image *Sombrero* was generated from a known parametric surface, using the following equation

$$Z(X, Y) = 0.5 \frac{\sin(r(X, Y))}{r(X, Y)} + 1.7, \quad r(X, Y) = \sqrt{(10X)^2 + (10Y)^2}. \quad (43)$$

The image was rendered using Eq. (5) at a size of 256×256 pixels, where the focal length was set $\mathbf{f} = 1$, the grid size was chosen to be $h_x = h_y = 1/200$ and the principal point was fixed at $c = (128, 128)^\top$. The second test image *Suzanne* was generated using the open-source software Blender [57]. In this context, the Z-buffer of the rendering path and the corresponding intrinsic parameters ($\mathbf{f} = 35$, $h_x = 1/16$, $h_y = 9/128$, $c = (256, 128)^\top$) were extracted and the final image was rendered at a size of 512×256 using Eq. (5) as before. The other two test images *Stanford Bunny* and *Dragon* have been computed likewise using 3-D models obtained from the Stanford 3D scanning repository [58]. For them a size of 256×256 pixels and the intrinsic parameters ($\mathbf{f} = 35$, $h_x = 1/8$, $h_y = 9/128$, $c = (128, 128)^\top$) were chosen. Finally, all images were saved as 8-bit grey-value images.

Let us finally comment on the selection of the parameters in our experiments. In order to keep the number of parameters low, we choose a preferred standard set of parameters for all the following experiments, unless otherwise stated: A down-sampling factor of $\eta = 0.8$ for the coarse-to-fine approach, $n = 10^6$ solver iterations on each coarse-to-fine level and a contrast parameter of $\lambda = 10^{-3}$. Moreover, the time step size τ provided in the different experiments always refers to the simplified explicit scheme. The time step size for the full explicit scheme is $\min(h_x^2, h_y^2)$ times smaller.

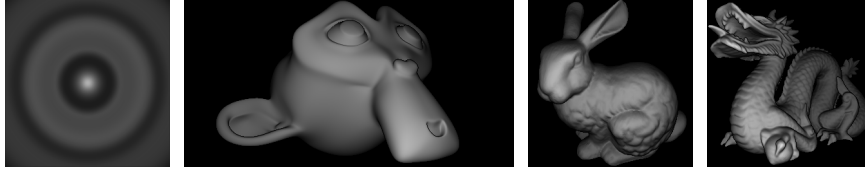


Fig. 2 Synthetic test images. **From left to right:** *Sombrero*, *Suzanne*, *Stanford Bunny* and *Dragon*.

Results on Synthetic Test Images. In our first experiment we evaluate the reconstruction quality of our novel approach. To this end, we applied our perspective SfS algorithm to all four of the previously discussed test images and compared the reprojected image and the reconstruction to the ground truth; see Fig. 3 and Fig. 4. Herein, the depth values are colour-coded in such a way that depth increases from red via green to blue. As one can see, both the reprojected image as well as the estimated depth values coincide very well with the ground truth. This is also confirmed by the corresponding surface error maps in Fig. 5. Indeed, only small differences for the *Stanford bunny* (right paw) and the *Dragon* (tail tip) are visible. As a consequence both error measures which are listed in Table 2 are very small. Moreover, one can see that the proposed subquadratic penaliser outperforms a quadratic smoothness term in most cases. Only for the *Sombrero* which has a rather smooth surface, the reconstruction error is smaller in the quadratic case.

Influence of the Regularisation. In our second experiment we investigate the influence of the regularisation on the quality of the reconstruction and its reprojection. To this end, we consider the *Sombrero* test image and vary the regularisation parameter α while the other parameters are kept fixed ($\tau = 0.001$, $n = 10^4$). The outcome is visualised in Fig. 6. While the reprojection related error measure (RIE) increases for a moderate amount of regularisation but is overall very low, the surface related error measure (RSE) decreases by almost a factor three (from 4.4×10^{-2} to 1.7×10^{-2}). This, however, is not surprising, since the computed surface typically exhibits some form of smoothness and thus benefits from a moderate amount of regularisation. Since the actual purpose of SfS is to find the correct surface, this shows that the regularisation may have an overall positive impact on the quality of the results.

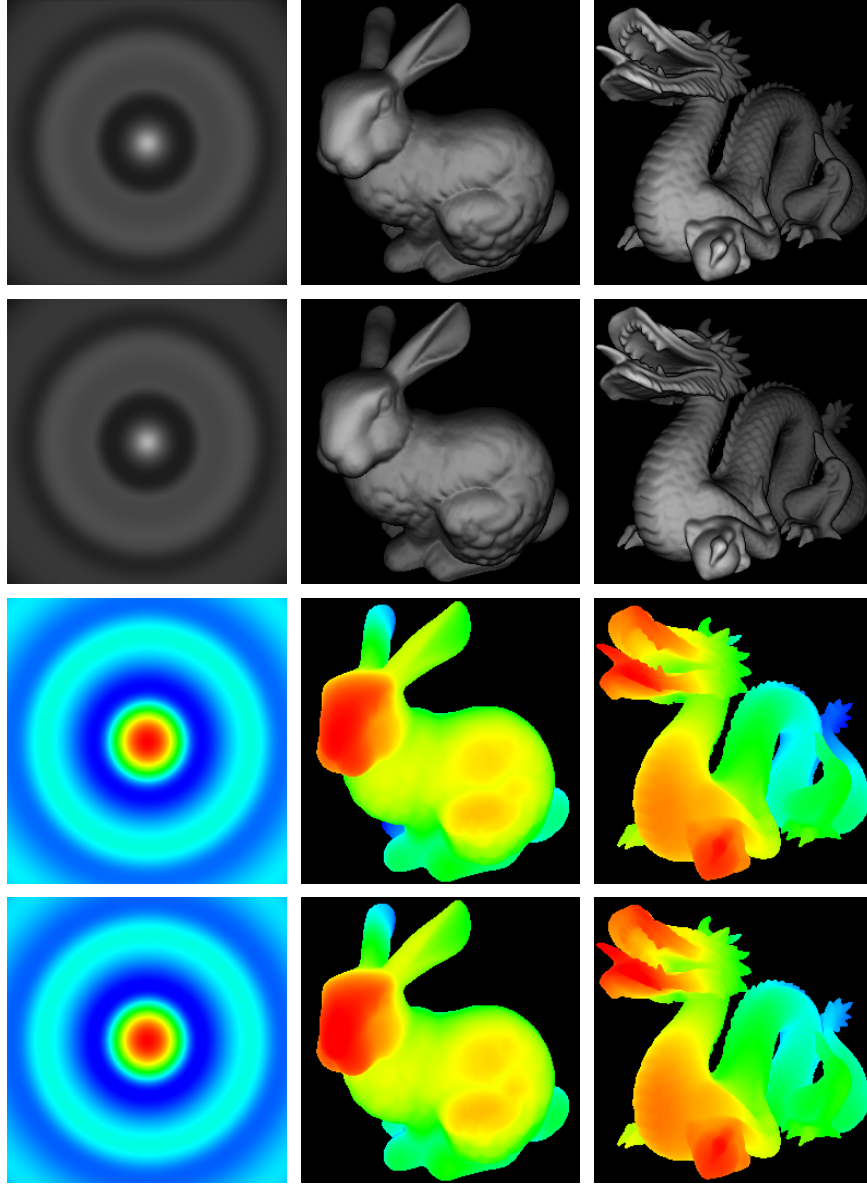


Fig. 3 First column, from top to bottom: Input image, reprojected image, ground truth depth, computed depth for the *Sombrero* test image ($\alpha = 7.5 \times 10^{-5}$, $\tau = 10^{-2}$, $n = 10^6$). **Second column:** Ditto for the *Stanford Bunny* test image ($\alpha = 7.5 \times 10^{-5}$, $\tau = 10^{-3}$, $n = 10^6$). **Third column:** Ditto for the *Dragon* test image ($\alpha = 7.5 \times 10^{-8}$, $\tau = 10^{-3}$, $n = 10^6$).

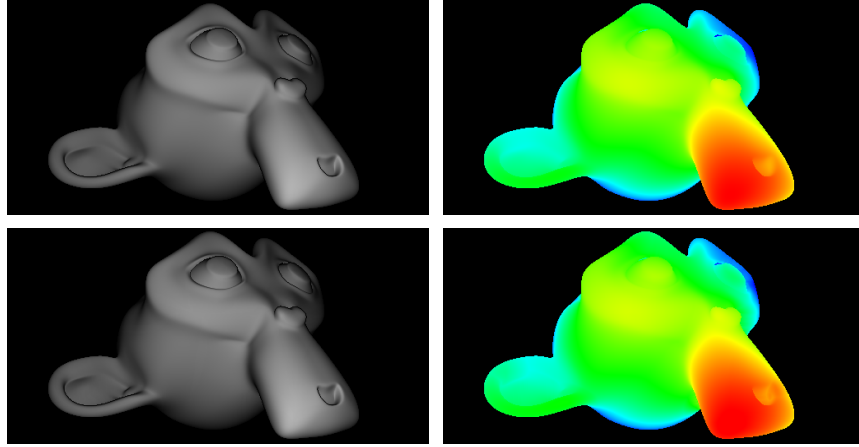


Fig. 4 First row, from left to right: Input image and ground truth depth of the *Suzanne* test image. Second row: Reprojected image and the computed depth ($\alpha = 10^{-7}$, $\tau = 10^{-3}$, $n = 10^6$).

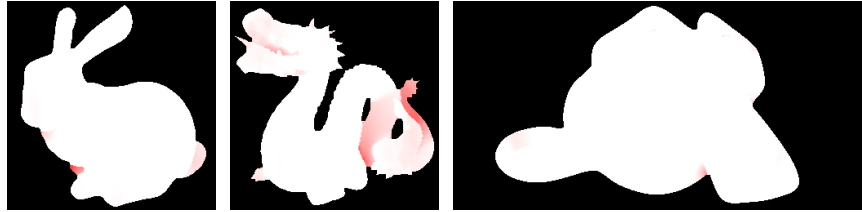


Fig. 5 Surface error maps. **From left to right:** *Stanford Bunny*, *Dragon* and *Suzanne*. Red denotes errors above 1 percent, where the intensity encodes the error magnitude. White denotes errors below 1 percent. The *Sombrero* is not shown, since the error is below 1 percent everywhere.

Table 2 Results for our approach with quadratic and subquadratic penaliser. Error measures are given in terms of the relative surface error (RSE) and the relative image error (RIE). Best results for each test image are highlighted boldface. Same parameters as in Fig. 3 and Fig. 4.

	quadratic		subquadratic		runtime
	RSE	RIE	RSE	RIE	
Sombrero	0.00208	0.00694	0.00318	0.00209	29113s
Stanford Bunny	0.00546	0.00015	0.00439	0.00007	23969s
Dragon	0.01376	0.00028	0.01376	0.00028	25350s
Suzanne	0.00392	0.00011	0.00251	0.00002	48395s

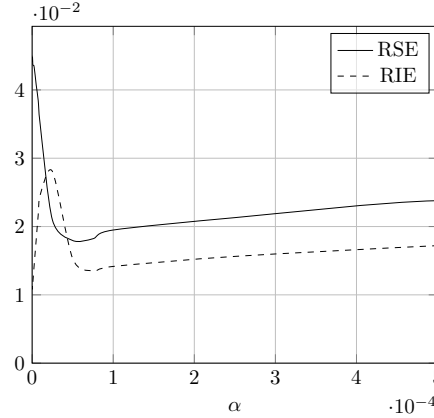


Fig. 6 Impact of the amount of regularisation on the reconstruction quality and the reprojection accuracy for the *Sombrero* test image.

Table 3 Impact of different initialisations on the reconstruction quality and reprojection accuracy for the *Stanford Bunny* ($\alpha = 7.5 \times 10^{-5}$, $\tau = 10^{-3}$, $n = 10^6$).

	initial error		after computation	
	RSE	RIE	RSE	RIE
plane ($z = 1$)	0.25804	1.63174	0.00439	0.00007
plane ($z = 10$)	6.41960	0.97373	0.00439	0.00007
proposed	0.37712	0.74363	0.00439	0.00007

Independence of the Initialisation. In our third experiment we analyse the dependency of our approach on the initialisation. To this end, we use the *Stanford Bunny* ($z \in [1, 2]$) and compare our initialisation on the coarsest scale of the proposed coarse-to-fine scheme (cf. Eq. (30)) with two other initialisations based on plain surfaces ($z = 1$, $z = 10$). The initial error and the outcome after $n = 10^6$ iterations are listed in Table 3. While the initial error for a good guess ($z = 1$) and a poor initialisation ($z = 10$) differs significantly, the quality of the reconstruction and the reprojection is identical after sufficiently many iterations. This also holds for our initialisation which can be computed from the input image without requiring a specific knowledge of the depth. That all initialisations converge to the same solution, however, is not surprising since the estimation is embedded in our coarse-to-fine scheme.

Comparison of Numerical Schemes. In our fourth experiment we compare the different numerical schemes proposed in Section 4: the full explicit scheme, the simplified explicit scheme and the alternating explicit scheme. In the first part of the

experiment we juxtapose the quality of the different numerical schemes for equal stopping times (iterations \times time step size). As one can see from the results in Table 4, the full explicit scheme clearly gives the best results in terms of reconstruction quality and reprojection accuracy. However, this comes at the expense of a significantly larger runtime, since more iterations are needed due to the time step restrictions discussed in Section 4. In fact the runtime is up to four orders of magnitude larger making the approach hardly feasible for larger image sizes. In the second part of the experiment we compared the numerical schemes for an equal number of iterations. From the results in Table 5 it becomes evident that in this case the simplified explicit scheme and in particular the alternating explicit scheme perform best in most cases in terms of reconstruction quality and reprojection accuracy. This demonstrates that it can be worthwhile to (partly) omit the terms that are added in the full explicit scheme since they slow down the convergence, but doing so does not necessarily compromise the quality.

Table 4 Comparison of different numerical schemes for equal stopping time $t = n \times \tau$. Results and runtimes refer to smaller versions of the four test images. Same parameters as in Fig. 3 and Fig. 4 except for n , which is given by $n = t/\tau$.

	alternating scheme		simplified scheme		full scheme	
test image	RSE	RIE	RSE	RIE	RSE	RIE
Small Sombrero (128 \times 128)	0.01823	0.01920 (runtime: 30s)	0.01820	0.02048 (runtime: 15s)	0.00785	0.00527 (runtime: 178021s)
Small Stanford Bunny (128 \times 128)	0.00659	0.00151 (runtime: 303s)	0.00667	0.00257 (runtime: 150s)	0.00576	0.00097 (runtime: 4278s)
Small Dragon (128 \times 128)	0.01667	0.00267 (runtime: 308s)	0.01673	0.00620 (runtime: 149s)	0.01526	0.00205 (runtime: 4304s)
Small Suzanne (128 \times 96)	0.00899	0.00514 (runtime: 223s)	0.01055	0.01909 (runtime: 111s)	0.01022	0.00203 (runtime: 2384s)

Reconstruction with Inpainting. In our fifth experiment we demonstrate the inpainting capabilities of the regularisation in combination with the confidence function c embedded in the data term. For this reason we created a pair of degraded *Stanford Bunny* test images together with the corresponding confidence functions, which are both depicted in Fig. 7. In addition, the computed depth values and the reprojected images are shown. One can see that in both cases the missing regions in the input image can hardly deteriorate the quality of the results since the smoothness term fills in the information from the neighbourhood. This is also reflected in the error measures given in Table 6. In case of the perforated version the surface error even remains the same compared to the result for the original version.

Table 5 Comparison of different numerical schemes for equal number of iterations. Results refer to the smaller versions of the four test images, see Tab. 4. The same parameters as in Fig. 3 and Fig. 4 have been used except for n , which is given by $n = 10^7$.

	alternating scheme		simplified scheme		full scheme	
test image	RSE	RIE	RSE	RIE	RSE	RIE
Small Sombrero	0.02357	0.00082	0.02392	0.00659	0.00358	0.00319
Small Stanford Bunny	0.00390	0.00001	0.00378	0.00004	0.00489	0.00047
Small Dragon	0.00572	0.00001	0.00562	0.00001	0.00964	0.00170
Small Suzanne	0.00319	0.00002	0.00320	0.00001	0.00505	0.00056

Table 6 Evaluation of inpainting properties for degraded versions of the *Stanford Bunny* test image. Same parameters as in Figure 7.

	perforated version (Fig. 7, top row)	sliced version (Fig. 7, bottom row)	original version (Fig. 2)
RSE	0.00439	0.00509	0.00439
RIE	0.00039	0.00249	0.00007

Comparison with a PDE-Based Approach. In our seventh experiment we compare the results of our variational method with the PDE-based approach of Vogel *et al.* [50] with Lambertian reflectance model. This essentially comes down to a comparison to the baseline method of Prados *et al.* [39] which is solved by Vogel *et al.* [50] as part of a Phong-based model using an efficient fast marching scheme [45]. In this experiment we consider two scenarios, that nicely demonstrate the advantages and shortcomings of the different types of methods: On the one hand, we use input images without noise, on the other hand, we added Gaussian noise of standard deviation $\sigma = 20$ before applying the two methods. The corresponding results are summarised in Tables 7 and 8, respectively. For the test images without noise both approaches give excellent results with errors among or below 1 percent of the solution. Thereby the approach of Vogel *et al.* gives slightly better results in terms of the relative surface error (RSE), while the variational approach gives better results in terms of the relative image error (RIE). From the viewpoint of the variational approach this can be explained as follows: While the data term penalises deviations from the photometric reprojection error and thus gives rather small RIE values, the regulariser and the coarse-to-fine scheme yield a moderate smoothing of the surface

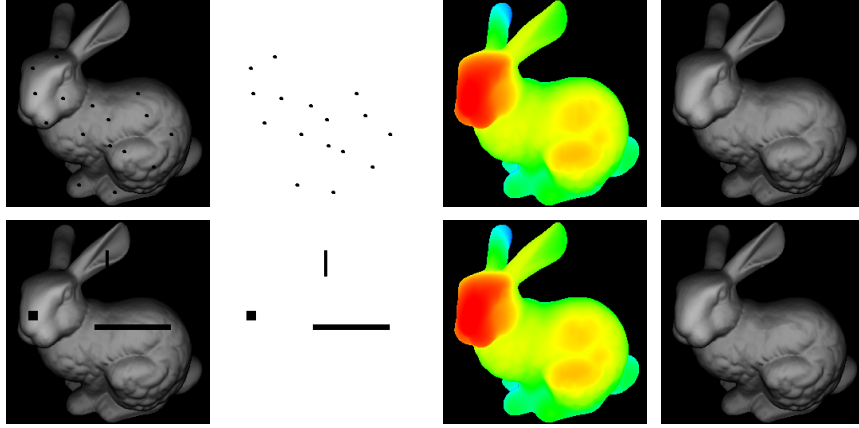


Fig. 7 First row, from left to right: Perforated version of the *Stanford Bunny* test image, corresponding confidence function c , computed depth values, reprojected image ($\alpha = 7.5 \times 10^{-5}$, $\tau = 10^{-3}$, $n = 10^6$). Second row: Ditto for the sliced version (same parameters).

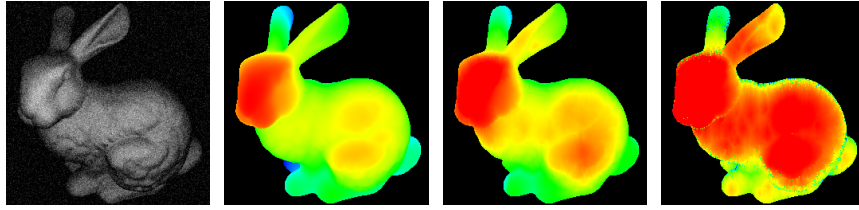


Fig. 8 From left to right: Noisy version of the *Stanford Bunny* (Gaussian noise with $\sigma = 20$), ground truth depth, computed depth using our variational approach ($\alpha = 1.0$, $\tau = 10^{-5}$, $n = 10^6$), computed depth using the PDE-based approach of Vogel *et al.* [50] with Lambertian model.

resulting in slightly higher RSE values. In the case of the noisy input images the findings are completely different. Here, the variational method can take advantage of both the regulariser and the independence of the initialisation. While a higher smoothness weight allows to obtain a smooth surface, the hierarchical initialisation via the coarse-to-fine scheme does not require to rely on noisy solutions at critical points as the PDE-based approach of Vogel *et al.* As a consequence, the resulting surface errors of 3 to 6 percent for our variational approach are significantly lower than those of the PDE-based model (11 to 20 percent). This can also be seen from the depth estimates for the *Stanford Bunny* depicted in Figure 8. Not surprisingly our findings are in full accordance with the observation in [29], in which the robustness of variational methods for perspective SfS has been investigated.

Results on Real-World Images. Finally, in order to evaluate our approach on real-world images, we used two images of faces provided by Prados [40]. According to Prados, these images have been taken with a cheap digital camera in a dark

Table 7 Comparison between our variational method and the PDE-based approach of Vogel *et al.* [50] with Lambertian reflectance model (= baseline model of Prados *et al.* [39]). Error measures are given in terms of the relative surface error (RSE) and the relative image error (RIE). Same parameters as in Fig. 3 and Fig. 4.

	Vogel <i>et al.</i> [50] (PDE-based approach)		our method (variational method)	
	RSE	RIE	RSE	RIE
Sombrero	0.00301	0.00495	0.00318	0.00209
Stanford Bunny	0.00266	0.00154	0.00439	0.00007
Dragon	0.00422	0.00255	0.01376	0.00028
Suzanne	0.00253	0.00082	0.00251	0.00002

Table 8 Performance under noise. Comparison between our variational method and the PDE-based approach of Vogel *et al.* [50] with Lambertian reflectance model (= baseline model of Prados *et al.* [39]). Gaussian noise of standard deviation $\sigma = 20$. Error measures are given in terms of the relative surface error (RSE) and the relative image error (RIE). The applied parameters are as follows: *Sombrero* ($\alpha = 0.1$, $\tau = 10^{-5}$, $n = 10^6$), *Stanford Bunny* ($\alpha = 1.0$, $\tau = 10^{-5}$, $n = 10^6$), *Dragon* ($\alpha = 1.0$, $\tau = 10^{-5}$, $n = 10^6$), *Suzanne* ($\alpha = 1.0$, $\tau = 5 \times 10^{-6}$, $n = 10^6$).

	Vogel <i>et al.</i> [50] (PDE-based approach)		Our method. (variational method)	
	RSE	RIE	RSE	RIE
Noisy Sombrero	0.19530	0.27254	0.05118	0.13239
Noisy Stanford Bunny	0.10973	0.17347	0.03235	0.15279
Noisy Dragon	0.12240	0.19409	0.05395	0.18767
Noisy Suzanne	0.12134	0.16783	0.01256	0.14302

place, where the scene is illuminated by the flash of the camera. The focal length is $f = 5.8\text{mm}$ and the grid size is approximately $h_x = h_y = 0.018\text{mm}$. The test images as well as additional images rendered from a new viewpoint using the computed depth are shown in Fig. 9. In both cases the results look quite realistic. One can also see how the depth values at the eyes have been inpainted in the reconstruction, since a manually defined confidence function was used to mask out those regions where the assumption of a Lambertian surface is violated.

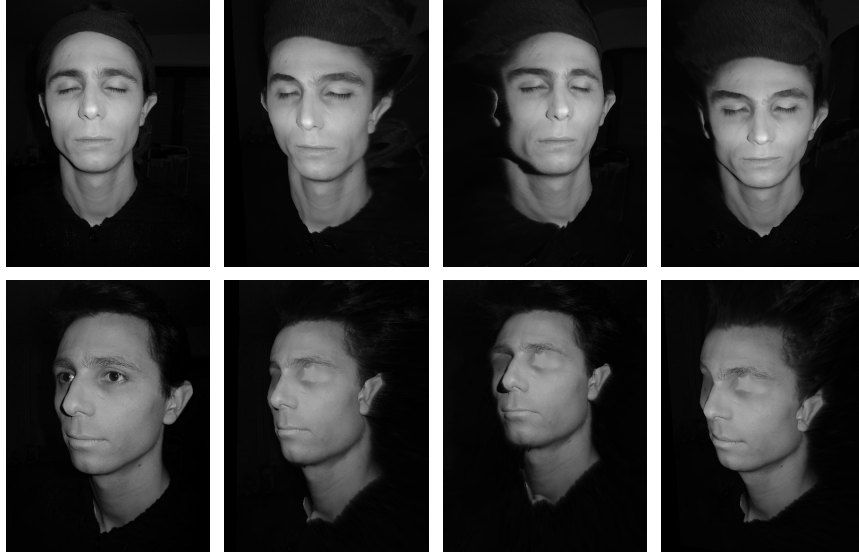


Fig. 9 First row, from left to right: Face with closed eyes, three images rendered from a new viewpoint using the estimated depth ($\alpha = 7.5 \times 10^{-5}$, $\tau = 5 \times 10^{-3}$, $n = 2 \times 10^5$). **Second row:** Ditto for the second test image ($\alpha = 7.5 \times 10^{-5}$, $\tau = 5 \times 10^{-3}$, $n = 2 \times 10^5$).

7 Conclusion

In this paper, we described a novel variational model for perspective shape from shading that not only has many desirable theoretical properties but also yields very convincing reconstruction results for synthetic and real-world input images, even in the presence of noise or other deteriorations in an input image. While the arising optimisation problem has turned out to be challenging, we have proposed an alternating explicit scheme embedded in a coarse-to-fine framework that is robust with respect to the initialisation and that allows reasonable computation times compared to a standard explicit scheme.

Besides the results that are documented via extensive experiments in this chapter, let us point out that we see a main contribution of our work in a different context, as we have laid the fundamental building block for a conceptually correct, working variational framework that can combine perspective shape from shading with other techniques from computer vision such as e.g. stereo vision. We aim to explore the arising possibilities in a future work.

Acknowledgements. This work has been partially funded by the Deutsche Forschungsgemeinschaft (DFG) as a joint project (BR 2245/3-1, BR 4372/1-1).

8 Appendix

Alternative Derivation of the Surface Normal. Instead of computing the derivatives with respect to the 2-D image coordinates x and y , one can also derive the surface normal in an alternative way that is often used in the literature, see e.g. [52]. The idea is to interpret the original surface in Eq. (4) as a function of the 3-D coordinates X , Y and $Z(X, Y)$

$$\mathcal{S}(X(\mathbf{x}, z), Y(\mathbf{x}, z), Z(X(\mathbf{x}, z), Y(\mathbf{x}, z))) = \begin{bmatrix} X(\mathbf{x}, z) \\ Y(\mathbf{x}, z) \\ Z(X(\mathbf{x}, z), Y(\mathbf{x}, z)) \end{bmatrix} := \begin{bmatrix} \frac{zx}{f} \\ \frac{zy}{f} \\ -z \end{bmatrix}. \quad (44)$$

Dropping the dependency of X , Y and $Z(X, Y)$ on \mathbf{x} , z and computing the partial derivatives with respect to X and Y via the chain rule

$$\frac{\partial X}{\partial Y} = \frac{\partial X}{\partial x} \frac{\partial x}{\partial Y}, \quad \frac{\partial Y}{\partial X} = \frac{\partial Y}{\partial y} \frac{\partial y}{\partial X}$$

then gives the tangent vectors to the surface

$$\mathcal{S}_X(\mathbf{x}, z) = \begin{bmatrix} 1 \\ \frac{z_y y}{z + z_x x} \\ -\frac{z_x f}{z + z_x x} \end{bmatrix}, \quad \mathcal{S}_Y(\mathbf{x}, z) = \begin{bmatrix} \frac{z_y x}{z + z_y y} \\ 1 \\ -\frac{z_y f}{z + z_y y} \end{bmatrix}. \quad (45)$$

After some computations we finally obtain the corresponding normal direction

$$\hat{\mathbf{n}}(\mathbf{x}) = \mathcal{S}_X(\mathbf{x}, z) \times \mathcal{S}_Y(\mathbf{x}, z) = \frac{f^2}{(z + z_x x)(z + z_y y)} \mathbf{n}(\mathbf{x}). \quad (46)$$

where $\mathbf{n}(\mathbf{x})$ is the normal direction from Eq. (8). As expected, both vectors only differ by scale, i.e. they have the same direction. Hence, the corresponding normalised vectors $\mathbf{n}/|\mathbf{n}|$ and $\hat{\mathbf{n}}/|\hat{\mathbf{n}}|$ are identical. While this alternative derivation was not used in our paper, it helps to clarify a common mistake in the literature that will be explained in the following.

Remark. Please note that, unlike in the orthographic case, the cross derivatives $\partial X/\partial Y$ and $\partial Y/\partial X$ do not vanish for the perspective model. Hence, using the orthographic derivation of the normal direction from Horn and Brooks [27]

$$\mathbf{n}_{\text{ortho}}(\mathbf{x}) = \frac{\partial}{\partial X} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \times \frac{\partial}{\partial Y} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ Z_X \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ Z_Y \end{bmatrix} = \begin{bmatrix} -Z_X \\ -Z_Y \\ 1 \end{bmatrix} \quad (47)$$

with zero cross derivatives and simply replacing the remaining partial derivatives Z_X and Z_Y by the corresponding expressions from (45) is *not completely correct* for the perspective case. Such an approach has for instance been proposed in [52, 55]. It actually mixes the orthographic and the perspective model and thus typically gives worse results in the case of strong perspective distortions. Moreover, apart from not being completely correct, this strategy also yields significantly more complex models that typically require auxiliary variables to be solved, see again e.g. [52, 55].

References

1. Abdelrahim, A.S.: Three-Dimensional Modeling of the Human Jaw/Teeth Using Optics and Statistics. PhD thesis, Department of Electrical and Computer Engineering, University of Louisville, Louisville, Kentucky, USA (2014)
2. Abdelrahim, A.S., Abdelrahman, M.A., Abdelmunim, H., Farag, A., Miller, M.: Novel image-based 3D reconstruction of the human jaw using shape from shading and feature descriptors. In: Proc. British Machine Vision Conference, 1–11 (2011)
3. Abdelrahim, A.S., Abdelmunim, H., Graham, J., Farag, A.: Novel variational approach for the perspective shape from shading problem using calibrated images. In: Proc. IEEE International Conference on Image Processing, 2563–2566 (2013)
4. Ahmed, A., Farag, A.: A new formulation for shape from shading for non-Lambertian surfaces. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1817–1824 (2006)
5. Barles, G.: Solutions de viscosité des équations de Hamilton-Jacobi. Mathématiques & Applications Vol. 17, Springer, 1994
6. Basha, T., Moses, Y., Kiryati, N.: Multi-view scene flow estimation: a view centered variational approach. International Journal of Computer Vision **101**(1), 6–21 (2012)
7. Bors, A.G., Hancock, E.R., Wilson, R.C.: Terrain analysis using radar shape-from-shading. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(8), 974–992 (2003)
8. Breuß, M., Cristiani, E., Durou, J.-D., Falcone, M., Vogel, O.: Numerical algorithms for perspective shape from shading. Kybernetika **46**(2), 207–225 (2010)
9. Breuß, M., Cristiani, E., Durou, J.-D., Falcone, M., Vogel, O.: Perspective shape from shading: ambiguity analysis and numerical approximations. SIAM Journal on Imaging Sciences **5**(1), 311–342 (2012)
10. Brooks, M.J., Horn, B.K.P.: Shape and source from shading. In: Proc. International Joint Conference in Artificial Intelligence, 932–936 (1985)
11. Brox, T., Bruhn, A., Papenberger, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Proc. European Conference on Computer Vision, LNCS Vol. 3024, 25–36 (2004)
12. Camilli, F., Prados, E.: Viscosity solution. In Katsushi Ikeuchi (Ed.), The Encyclopedia of Computer Vision, Springer, 2014
13. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M.: Deterministic edge-preserving regularization in computed imaging. IEEE Transactions on Image Processing **6**(2), 298–311 (1997)
14. Crandall, M.G., Lions, P.-L.: Viscosity solutions of Hamilton-Jacobi equations. Transactions of the American Mathematical Society **277**(1), 1–42 (1983)
15. Courant, R., Hilbert, D.: Methods of Mathematical Physics. Interscience Publishers, Inc., New York, NY (1953)
16. Courteille, F., Crouzil, A., Durou, J.-D., Gurdjos, P.: Towards shape from shading under realistic photographic conditions. In: Proc. IEEE International Conference on Pattern Recognition, 277–280 (2004)

17. Courteille, F., Crouzil, A., Durou, J.-D., Gurdjos, P.: 3D-spline reconstruction using shape from shading: spline from shading. *Image and Vision Computing* **26**(4), 466–479 (2008)
18. Demetz, O., Stoll, M., Volz, S., Weickert, J., Bruhn, A.: Learning brightness transfer functions for the joint recovery of illumination changes and optical flow. In: *Proc. European Conference on Computer Vision*, LNCS Vol. 8689, 455–471 (2014)
19. Diggelen, J.V.: A photometric investigation of the slopes and heights of the ranges of hills in the maria of the moon. *Bulletin of the Astronomical Institutes of the Netherlands* **XI**(423), 283–289 (1951)
20. Durou, J.-D., Falcone, M., Sagona, M.: Numerical methods for shape-from-shading: a new survey with benchmarks. *Computer Vision and Image Understanding* **109**(1), 22–43 (2008)
21. Estellers, V., Thiran, J.-P., Gabrani, M.: Surface reconstruction from microscopic images in optical lithography. *IEEE Transactions on Image Processing* **23**(8), 3560–3573 (2014)
22. Frankot, R.T., Chellappa, R.: A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**(4), 439–451 (1988)
23. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2004)
24. Horn, B.K.P.: *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, Department of Electrical Engineering, MIT, Cambridge, Massachusetts, USA (1970)
25. Horn, B.K.P.: *Robot Vision*. MIT Press (1986)
26. Horn, B.K.P., Brooks, M.J.: The Variational Approach to Shape from Shading. *Computer Vision, Graphics, and Image Processing* **33**, 174–208 (1986)
27. Horn, B.K.P., Brooks, M.J.: *Shape from Shading*. Artificial Intelligence Series, MIT Press (1989)
28. Ikeuchi, K., Horn, B.K.P.: Numerical shape from shading and occluding boundaries. *Artificial Intelligence* **17**(1–3), 141–184 (1981)
29. Ju, Y. C., Breuß, M., Bruhn, A.: Variational perspective shape from shading. In: *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, LNCS Vol. 9087, 538–550 (2015).
30. Kimmel, R., Siddiqi, K., Kimia, B.B., Bruckstein, A.M.: Shape from shading: level set propagation and viscosity solutions. *International Journal of Computer Vision* **16**, 107–133 (1995)
31. Lysaker, M., Lundervold, A., Tai, X.-C.: Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing* **12**(12), 1057–1590 (2003)
32. Mecca, R., Falcone, M.: Uniqueness and approximation of a photometric shape-from-shading model. *SIAM Journal on Imaging Science* **6**, 616–659 (2013)
33. Mecca, R., Wetzler, A., Bruckstein, A.M., Kimmel, R.: Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences* **7**(4), 2732–2770 (2014)
34. Okatani, T., Deguchi, K.: Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer Vision and Image Understanding* **66** (1997) 119–131
35. Oliensis, J.: Shape from shading as a partially well-constrained problem. *Computer Vision, Graphics, and Image Processing: Image Understanding* **54**(2), 163–183 (1991)
36. Oren, M., Nayar, S.: Generalization of the Lambertian model and implications for machine vision. *International Journal of Computer Vision* **14**(3), 227–251 (1995)
37. Phong, B.T.: Illumination for computer-generated pictures. *Communications of the ACM* **18**(6), 311–317 (1975)
38. Prados, E., Faugeras, O.: “Perspective shape from shading” and viscosity solutions. In: *Proc. IEEE International Conference Computer Vision*, 826–831 (2003)
39. Prados, E., Faugeras, O.: Shape from shading: a well-posed problem? In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 870–877 (2005)
40. Prados, E., Camilli, F., Faugeras, O.: A unifying and rigorous shape from shading method adapted to realistic data and applications. *Journal of Mathematical Imaging and Vision* **25**(3), 307–328 (2006)

41. Rindfleisch, T.: Photometric method for lunar topography. *Photogrammetric Engineering* **32**(2), 262–277 (1966)
42. Robert, L., Deriche, R.: Dense depth map reconstruction: a minimization and regularization approach which preserves discontinuities. In: *Proc. European Conference on Computer Vision*, LNCS Vol. 1064, 439–451 (1996)
43. Rouy, E., Tourin, A.: A viscosity solution approach to shape-from-shading. *SIAM Journal on Numerical Analysis* **29**(3), 867–884 (1992)
44. Schmaltz, C., Peter, P., Mainberger, M., Eberl, F., Weickert, J., Bruhn, A.: Understanding, optimising, and extending data compression with anisotropic diffusion. *International Journal of Computer Vision* **108**(3), 222–240 (2014)
45. Sethian, J.: *Level Set Methods and Fast Marching Methods*. Cambridge University Press (1999)
46. Tankus, A., Sochen, N., Yeshurun, Y.: A new perspective [on] shape-from-shading. In: *Proc. IEEE International Conference on Computer Vision*, 862–869 (2003)
47. Tankus, A., Sochen, N., Yeshurun, Y.: Shape-from-shading under perspective projection. *International Journal of Computer Vision* **63**(1), 21–43 (2005)
48. Vogel, O., Breuß, M., Weickert, J.: Perspective shape from shading with non-Lambertian reflectance. In: *Proc. German Conference on Pattern Recognition*, LNCS Vol. 5096, 517–526 (2008)
49. Vogel, O., Bruhn, A., Weickert, J., Didas, S.: Direct shape-from-shading with adaptive higher order regularisation. In: *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, LNCS Vol. 4485, 871–882 (2007)
50. Vogel, O., Breuß, M., Leichtweis, T., Weickert, J.: Fast shape from shading for Phong-type surfaces. In: *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, LNCS Vol. 5567, 733–744 (2009)
51. Wang, G.H., Han, J.Q., Zhang, X.M.: Three-dimensional reconstruction of endoscope images by a fast shape from shading method. *Measurement Science and Technology* **20**(12), (2009)
52. Wu, C., Narasimhan, S., Jaramaz, B.: A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *International Journal of Computer Vision* **86**, 211–228 (2010)
53. Yuen, S.Y., Tsui, Y.Y., Chow, C.K.: A fast marching formulation of perspective shape from shading under frontal illumination. *Pattern Recognition Letters* **28**(7), 806–824 (2007)
54. Zhang, R., Tsai, P.-S., Cryer, J.E., Shah, M.: Shape from shading: a Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(8), 690–706 (1999)
55. Zhang, L., Yip, A. M., Tan, C. T.: Shape from shading based on Lax-Friedrichs fast sweeping and regularization techniques with applications to document image restoration. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1–8 (2007)
56. Zhang, L., Yip, A. M., Tan, C. T.: A restoration framework for correcting photometric and geometric distortions in camera-based document images. In: *Proc. IEEE International Conference on Computer Vision*, 1–8 (2007)
57. <http://www.blender.org> (Last visited on 05/05/2015).
58. The Stanford 3D Scanning Repository, <http://graphics.stanford.edu/data/3Dscanrep/> (Last visited on 05/05/2015).